State University of New York at Buffalo
Department of Computer Science

(9)

ON THE STATIONARITY OF MULTIPLE AUTOREGRESSIVE APPROXIMANTS: THEORY AND ALGORITHMS*

by

H. Joseph Newton and Marcello Pagano

Statistical Science Division
State University of New York at Buffalo

TECHNICAL REPORT NO.: 43

August, 1976

* Research supported in part by the Office of Naval Research

ON THE STATIONARITY OF MULTIPLE AUTOREGRESSIVE APPROXIMANTS: THEORY AND ALGORITHMS

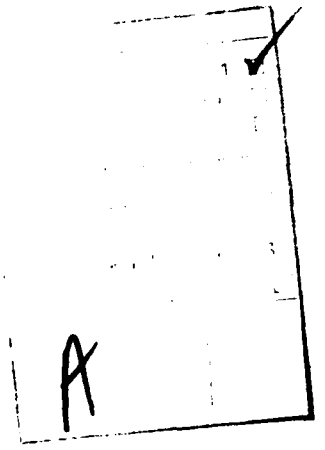
by

H. Joseph Newton and Marcello Pagano

Statistical Science Division
State University of New York at Buffalo

ABSTRACT

Numerical methods are presented for determining the stability of a multiple autoregressive scheme. The question of stability is important in prediction theory, but, furthermore, the methods for determining stability are such that they yield information which is important when fitting time series models to data.



1. Introduction

Let $\{\underset{\sim}{X}(t), t\epsilon Z\}$, Z the set of all integers, be a zero mean, d-dimensional, covariance stationary time series with absolutely summable autocovariance function $R(v) = E(\underset{\sim}{X}(t)\underset{\sim}{X}^T(t+v))$, $v\epsilon Z$. A^T denotes the transpose of the matrix A. Then the spectral density matrix,

$$f(\omega) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} R(v)\, e^{-iv\omega}, \qquad \omega\epsilon[-\pi, \pi]$$

and the $R(\cdot)$, are Fourier pairs,

$$R(v) = \int_{-\pi}^{\pi} f(\omega)\, e^{iv\omega}\, d\omega, \quad v\epsilon Z .$$

Formally define the inverse spectral density matrix by

$$fi(\omega) = \frac{1}{4\pi^2} f^{-*}(\omega), \qquad \omega\epsilon[-\pi, \pi],$$

where A^* denotes the complex conjugate transpose of the matrix A. If there exist positive finite constants λ_1, λ_2 such that

$$\lambda_1 I_d \le f(\omega) \le \lambda_2 I_d , \tag{1.1}$$

where I_d denotes the d-dimensional identity matrix, then $fi(\omega)$ exists and has integrable components and we can define the covinverse matrices

$$Ri(v) = \int_{-\pi}^{\pi} fi(\omega)\, e^{iv\omega}\, d\omega, \quad v\epsilon Z .$$

Further, (Masani (1966)) there exists an infinite order autoregressive representation for $\{\underset{\sim}{X}(t), t\epsilon Z\}$

$$\sum_{j=0}^{\infty} A_{\infty}(j)\underset{\sim}{X}(t-j) = \underset{\sim}{\varepsilon}(t) , \tag{1.2}$$

where

$$\underset{\sim}{\varepsilon}(t) = \underset{\sim}{X}(t) - E(\underset{\sim}{X}(t) \,|\, \underset{\sim}{X}(t-1), \underset{\sim}{X}(t-2), \ldots), t\epsilon Z ,$$

is the infinite memory one-step ahead prediction error and is a multiple white noise series, with positive definite covariance matrix Σ_{∞},

$$E(\underset{\sim}{\varepsilon}(t)\underset{\sim}{\varepsilon}^{T}(t+v)) = \delta_{v,0}\,\Sigma_{\infty} ,$$

where $\delta_{j,k} = 1$, if $j = k$, and zero otherwise. Defining L to be the backward lag operator, $L^{j}\underset{\sim}{X}(t) = \underset{\sim}{X}(t-j)$, $j\epsilon Z$, we can write (1.2) as

$$G_{\infty}(L)\underset{\sim}{X}(t) = \underset{\sim}{\varepsilon}(t) ,$$

where the complex valued matrix polynomial $G_{\infty}(\cdot)$ is defined by

$$G_{\infty}(z) = \sum_{j=0}^{\infty} A_{\infty}(j)z^{j} .$$

The representation (1.2) affords a representation of $f(\cdot)$:

$$f_\infty(\omega) = \frac{1}{2\pi} G_\infty^{-1}(e^{i\omega}) \Sigma_\infty G_\infty^{-*}(e^{i\omega}) , \quad \omega \in [-\pi, \pi] .$$

Definition of Autoregressive Approximant

Let

$$f_p(\omega) = \frac{1}{2\pi} G_p^{-1}(e^{i\omega}) \Sigma_p G_p^{-*}(e^{i\omega}) , \quad \omega \in [-\pi, \pi]$$

where

$$G_p(z) = \sum_{j=0}^{p} A_p(j) z^j ,$$

and $A_p(1), \ldots, A_p(p), \Sigma_p$ are solutions to the Yule-Walker equations,

$$\sum_{j=0}^{p} A_p(j) R(j - v) = \delta_{v,0} \Sigma_p , \qquad v = 0, \ldots, p . \tag{1.3}$$

We call $f_p(\cdot)$ the p^{th} order autoregressive approximant to $f(\cdot)$ (Parzen (1975)). The use of the term autoregressive is natural since the p-memory linear prediction error, $\underset{\sim}{\varepsilon}^{(p)}$, defined by

$$\underset{\sim}{\varepsilon}^{(p)}(t) = \underset{\sim}{X}(t) - E(\underset{\sim}{X}(t) \,|\, \underset{\sim}{X}(t-1), \ldots, \underset{\sim}{X}(t-p))$$

is given by

$$\underset{\sim}{\varepsilon}^{(p)}(t) = \sum_{j=0}^{p} A_p(j)\underset{\sim}{X}(t - j) , \qquad A_p(0) \equiv I_d , \tag{1.4}$$

where $A_p(1),\ldots,A_p(p)$, and $\Sigma_p = E(\underset{\sim}{\varepsilon}^{(p)}(t)\underset{\sim}{\varepsilon}^{(p)T}(t))$

are obtained from (1.3).

Looking at autoregressive approximants from a prediction theory point of view we see that for stability of predictions, it is desirable that we be able to write $\underset{\sim}{X}(t)$ as a linear combination only of past and present $\underset{\sim}{\varepsilon}^{(p)}(t)$:

$$\underset{\sim}{X}(t) = \sum_{k=0}^{\infty} B_{p,\infty}(k)\underset{\sim}{\varepsilon}^{(p)}(t - k) , \qquad t\varepsilon Z . \tag{1.5}$$

Definition of Stationary Autoregressive Approximant

An autoregressive approximant of the form (1.4) which has a representation, as a limit in the mean, of the form (1.5) is said to be stationary.

The use of autoregressive approximants for $d = 1$ has become quite widespread (Parzen (1974), Akaike (1971)). Parzen (1976) has recently extended his method of determining the order of autoregressive approximants to the multiple case. In Section 2 we develop theory and algorithms for determining whether a given set of $A(1),\ldots,A(p)$, and Σ determine a stationary autoregressive approximant. This theory allows us to obtain two further results:

(a) Autoregressive approximants for all orders from a covariance stationary time series whose spectral density is bounded above and below by positive finite constants are stationary.

(b) Algorithms for determining the autocovariances $R(0),\ldots,R(p)$ corresponding to the $A(\cdot)$ and Σ of a stationary autoregressive approximant. These algorithms have several applications in multiple time series analysis, two of which are given in Section 3.

Finally, in Section 4 we give a numerical example of the algorithms developed in Section 2.

2. The Algorithms

Given matrices $A(1),\ldots,A(p)$ and Σ, define the autoregressive process $\{\underset{\sim}{X}(t), t\epsilon Z\}$ by

$$\sum_{j=0}^{p} A(j)\underset{\sim}{X}(t-j) = \underset{\sim}{\varepsilon}(t) \quad , \tag{2.1}$$

where $\{\underset{\sim}{\varepsilon}(t), t\epsilon Z\}$ is a white noise process having covariance matrix Σ. We develope algorithms for determining whether there exists a stationary solution to (2.1) in terms of only present and past $\underset{\sim}{\varepsilon}(\cdot)$.

Lemma 1: (Hannan (1970, p. 14))

If $\{\underset{\sim}{\varepsilon}(t), t\epsilon Z\}$ is a white noise series, then given p initial values, $\underset{\sim}{X}(-p),\ldots,\underset{\sim}{X}(-1)$, the solution to (2.1) is uniquely defined. If all zeros of the determinant $|G(z)|$ are outside the unit circle, the solution is of the form

$$\underset{\sim}{X}(t) = \sum_{k=0}^{\infty} B(k)\underset{\sim}{\varepsilon}(t-k), \qquad t\epsilon Z,$$

where the elements $B_{ij}(k)$ of the $B(k)$ converge to zero exponentially as $k \to \infty$.

The $B(\cdot)$ can be obtained by comparing coefficients of z^j in

$$G^{-1}(z) = H(z) = \sum_{k=0}^{\infty} B(k)z^k .$$

Method 1 (via the Schur matrix)

Method 1 consists of determining whether the zeros of $|G(z)|$ are outside the unit circle. Writing $|G(z)|$ as a polynomial of degree pd,

$$|G(z)| = \sum_{j=0}^{pd} \gamma(j)z^j ,$$

(see Robinson (1967), p. 162 for example) we can implement

Theorem on the Schur Matrix (Schur (1917))

The zeros of the polynomial $g(z) = \sum_{j=0}^{n} a(j)z^j$ are outside the unit circle if and only if the matrix S defined by $S = (S_{jk})$,

$$S_{jk} = \sum_{\ell=0}^{v} [a(j-1-\ell)a(k-1-\ell) - a(n+1+\ell-j)a(n+1+\ell-k)] ,$$

$j, k = 1, \ldots, n$, $v = \min(j-1, k-1)$, is positive definite. The matrix S is called the Schur matrix of the polynomial $g(\cdot)$. It is particularly useful when $d = 1$, since then (Pagano (1973)), if the zeros of $|G(z)|$ are all outside the unit circle,

$$S^{-1} = \frac{1}{\sigma^2}\Gamma_p , \qquad (2.2)$$

where Γ_p is the $p \times p$ Toeplity matrix having $R(j - k)$ in the j^{th} row and k^{th} column, while $\sigma^2 = E(\epsilon^2(t))$. Unfortunately there appears to be no vector analogue of (2. 2).

To decide whether S is positive definite one may calculate its modified Cholesky decomposition (see Wilkinson (1967)). Then S is positive definite if, and only if, the diagonal elements of the diagonal matrix in its modified Cholesky decomposition are positive.

As mentioned above, method 1 for $d = 1$ produces the autocovariances of $\underset{\sim}{X}(\cdot)$, while for $d > 1$ it does not. The methods given below do give the $R(\cdot)$. They are based on the following lemmas.

Lemma 2 (Hannan (1970), p. 19)

Let $R(v) = E(\underset{\sim}{X}(t)\underset{\sim}{X}^T(t + v))$, $v \epsilon Z$. If the zeros of $|G(z)|$ are all outside the unit circle, then the $R(\cdot)$ sequence satisfies

$$\sum_{j=0}^{p} A(j)R(j - v) = \delta_{v,0}\Sigma, \quad v \geq 0. \tag{2. 3}$$

Lemma 3

If Σ is positive definite, then the zeros of $|G(z)|$ are all outside the unit circle if, and only if, Γ_p is positive definite, where Γ_p is the $p \times p$ block Toeplitz matrix having $R(j - k)$ in the j^{th} row and k^{th} column of blocks, i. e., $\Gamma_p = \text{Toepl}(R(0), \ldots, R(1 - p))$.

Proof

Hannan (1970), p. 329, shows sufficiency, while Whittle (1963) shows necessity.

Theorem: Bounded Spectra Guarantees Stationary Approximants

If $\{\underset{\sim}{X}(t), t\epsilon Z\}$ is a covariance stationary time series whose spectral density matrix satisfies (1.1), then the autoregressive approximants for $p \geq 1$ defined by (1.4) are stationary.

Proof: . . .

We show that Σ_p and Γ_p are positive definite for $p \geq 1$, thus ensuring that the zeros of $\left| \sum_{j=0}^{p} A_p(j) z^j \right|$ are outside the unit circle.

Define the nonnull, pd-vector $\underset{\sim}{\ell}^T = (\underset{\sim}{\ell}_1^T : \ldots : \underset{\sim}{\ell}_p^T)$, where $\underset{\sim}{\ell}_j$ is a d-vector. Then Γ_p is positive definite if, and only if, $\underset{\sim}{\ell}^T \Gamma_p \underset{\sim}{\ell} > 0$ for all such $\underset{\sim}{\ell}$. Now

$$\underset{\sim}{\ell}^T \Gamma_p \underset{\sim}{\ell} = \sum_{j,k=1}^{p} \underset{\sim}{\ell}_j^T R(j-k) \underset{\sim}{\ell}_k$$

$$= \sum_{j,k=1}^{p} \underset{\sim}{\ell}_j^T \left[\int_{-\pi}^{\pi} f(\omega)\, e^{i(j-k)\omega}\, d\omega \right] \underset{\sim}{\ell}_k$$

$$= \int_{-\pi}^{\pi} \underset{\sim}{C}^*(\omega) f(\omega) \underset{\sim}{C}(\omega)\, d\omega ,$$

where $\underset{\sim}{C}(\omega) = \sum_{j=1}^{p} \underset{\sim}{\ell}_j e^{ij\omega}$ is nonnull. Now (1.1) implies

$$\underset{\sim}{\ell}^T \Gamma_p \underset{\sim}{\ell} \geq \lambda_1 \int_{-\pi}^{\pi} \underset{\sim}{C}^*(\omega)\underset{\sim}{C}(\omega)\, d\omega$$

$$= \lambda_1 \sum_{j,k=1}^{p} \underset{\sim}{\ell}_j^T \underset{\sim}{\ell}_k \int_{-\pi}^{\pi} e^{i(j-k)\omega}\, d\omega$$

$$= \lambda_1 \sum_{j=1}^{p} \underset{\sim}{\ell}_j^T \underset{\sim}{\ell}_j > 0 \quad .$$

Thus we can write the Yule-Walker equations (1, 3) as

$$\Gamma_p A_p = -R_p \,,$$

$$\Sigma_p = R(0) - R_p^T \Gamma_p^{-1} R_p \,,$$

where $A_p = (A_p(1) \vdots \ldots \vdots A_p(p))^T$, and $R_p = (R^T(1) \vdots \ldots \vdots R^T(p))^T$. Now Γ_p and Γ_{p+1} are positive definite, and by the clockwise rule of matrix determinants,

$$|\Gamma_{p+1}| = \begin{bmatrix} R(0) & R_p^T \\ R_p & \Gamma_p \end{bmatrix}$$

$$= |\Gamma_p| \; |R(0) - R_p^T \Gamma_p^{-1} R_p|$$

$$= |\Gamma_p| \; |\Sigma_p| \quad .$$

Since Σ_p is a covariance matrix, the proof is complete.

Method 2 (via the Yule-Walker equations)

Method 2 consists of solving (2, 3) for the $R(\cdot)$ and then checking the resulting Γ_p for positive definiteness. The transformation from the $A(\cdot)$ to the $R(\cdot)$ is not obvious (see Quenoville (1957)). The symmetry of $R(0)$ and the noncommutativity of the $R(\cdot)$ and $A(\cdot)$ seems to rule out a solution of (2, 3) in terms of the matrices themselves. The matrix notation in (2, 3) disguises the fact that there are in fact $(p + 1)d^2$ scalar equations in the first $p + 1$ matrix equations. Further, the symmetry of $R(0)$ and Σ actually leaves $s = pd^2 + d(d + 1)/2$ distinct scalar knowns and s unknowns in these equations, namely

$$\sum_{j=0}^{p} \sum_{\ell=1}^{d} A_{m\ell}(j) R_{\ell n}(j - v) = \delta_{v,0} \Sigma_{mn}, \tag{2.4}$$

for $v = 0, \ldots, p$; $m, n = 1, \ldots, d$.

Define the vec operator, vec (C) , on the $(n \times m)$ matrix $C = (\underset{\sim}{C}_1, \ldots, \underset{\sim}{C}_m)$ by

$$\text{vec } (C) = (\underset{\sim}{C}_1^T \vdots \ldots \vdots \underset{\sim}{C}_m^T)^T ,$$

and the lvec operator, lvec (C) , as the vec operator on the lower triangular portion of C . Then define the $(s \times 1)$ vectors $\underset{\sim}{r}$ and $\underset{\sim}{g}$ by

$$\underset{\sim}{r} = (\underset{\sim}{r}^T(0) \vdots \ldots \vdots r^T(p))^T$$

$$\underset{\sim}{g} = (\underset{\sim}{g}^T(0) \vdots \ldots \vdots \underset{\sim}{g}^T(p))^T ,$$

where

$$\underset{\sim}{r}(j) = \begin{cases} \text{lvec } (R(0)) \ , & j = 0 \\ \text{vec } (R(j)) \ , & j = 1, \ldots, p \end{cases}$$

$$\underset{\sim}{g}(j) = \begin{cases} \text{lvec } (\Sigma) \ , & j = 0 \\ \underset{\sim}{0}_{d^2} \ , & j = 1, \ldots, p \ , \end{cases}$$

where $\underset{\sim}{0}_n$ denotes the n-vector of zeros. Then the equations (2.4) can be written

$$W\underset{\sim}{r} = \underset{\sim}{g} \ , \tag{2.5}$$

where W is an $(s \times s)$ matrix determined as follows.

Since $R(-v) = R^T(v)$, we have

$$R_{jk}(v) = \begin{cases} R_{jk}(v) \ , & v = 1, \ldots, p \\ R_{kj}(-v) \ , & v = -p, \ldots, -1 \\ R_{m_1 m_2}(0) \ , & v = 0 \ , \end{cases} \tag{2.6}$$

where $m_1 = \max(j, k)$, $m_2 = \min(j, k)$. The row number of $R_{jk}(v)$ in $\underset{\sim}{r}$ is given by

$$i_{jk}(v) = \begin{cases} (k-1)d + j - k(k-1)/2 \ , & v = 0 \\ d(d+1)/2 + (v-1)d^2 + (k-1)d + j \ , & v = 1, \ldots, p \\ & j, k = 1, \ldots, d \ . \end{cases}$$

Thus to determine W :

(a) Set W initially equal to zero

(b) For fixed m , n , v , j , ℓ in (2.4) :

(i) The row number of W corresponding to m , n , v is given by $t_1 = i_{mn}(v)$.

(ii) Find by (2.6) the indices ℓ' , n' , u' of the representation $R_{\ell'n'}(u')$ of $R_{\ell n}(u)$, $u = j - v$, for nonnegative u' .

(iii) Add $A_{m\ell}(j)$ to $W_{t_1 t_2}$ where $t_2 = i_{\ell'n'}(u')$.

If (2.5) cannot be solved uniquely for $\underset{\sim}{r}$, then the process is not stationary. If (2.5) can be solved, it is not difficult to reorder the elements of $\underset{\sim}{r}$ to form $R(p),\ldots,R(p)$, and Γ_p , and to check Γ_p for positive definiteness.

Note that it is possible to solve (2.5) uniquely for $\underset{\sim}{r}$ while Γ_p is not positive definite. For example, let $p = d = 1$. Then (2.5) becomes

$$\begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix} \quad \begin{pmatrix} R(0) \\ R(1) \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \end{pmatrix} .$$

If $|\alpha| > 1$ and $\sigma^2 > 0$, then $\Gamma_1 = R(0) = \sigma^2/(1 - \alpha^2) < 0$.

Method 3 (iterative)

If d is large, the value of s may prohibit the use of Method 2 unless one uses an equation solver to solve (2.5) which does not require storage of the entire W matrix. Thus we propose an iterative method for calculating Γ_p which requires much less storage than the Yule-Walker equation method.

By lemma 1, if the zeros of $|G(z)|$ are all outside the unit circle, we can write

$$G^{-1}(z) = H(z) = \sum_{k=0}^{\infty} B(k)z^k , \tag{2.7}$$

where the $(\ell, m)^{th}$ element of $B(k)$ converges exponentially to zero as $k \to \infty$, i.e., there exist constants $C \neq 0$, ρ such that for k sufficiently large,

$$|B_{\ell m}(k)| \leq C\rho^k , \qquad |C| < \infty , \ |\rho| < 1 .$$

Thus we can obtain $R(0), \ldots, R(p)$ by

$$R(v) = \sum_{k=0}^{\infty} B(k)\Sigma B^T(k+v) , \qquad v = 0, \ldots, p . \tag{2.8}$$

The $B(\cdot)$ can be obtained by solving the block triangular system of equations

$$G(z)H(z) = I_d . \tag{2.9}$$

Equating coefficients of z^j in (2.9), we have the difference equation with $B(0) = I_d$,

$$\sum_{j=0}^{\min(\ell, p)} A(j)B(\ell - j) = 0 \quad , \qquad \ell > 0 \, . \tag{2.10}$$

$R(v)$ can be approximated by

$$R_M(v) = \sum_{k=0}^{M-v} B(k)\Sigma B^T(k + v) \quad , \qquad v = 0, \ldots, p \, . \tag{2.11}$$

Then we have the following.

Convergence Theorem for $R_M(\cdot)$

Let $\|C\|$ denote a norm of an $(n \times n)$ matrix $C = (C_{jk})$. Then a necessary condition for $A(1), \ldots, A(p)$, and Σ to define a stationary autoregressive approximant of order p is that

$$\|R(v) - R_M(v)\| = O(\gamma^v) \quad , \qquad |\gamma| < 1 \quad ,$$

uniformly in $v = 0, \ldots, p$.

Proof:

From (2.8) and (2.11) we have, if the zeros of $|G(z)|$ are all outside the unit circle,

$$\|R(v) - R_M(v)\| = \| \sum_{k=M-v+1}^{\infty} B(k)\Sigma B^T(k+v) \|$$

$$\leq \sum_{k=M-p+1}^{\infty} \|B(k)\| \, \|\Sigma\| \, \|B^T(k+v)\|$$

$$\leq C_1 \sum_{k=M-p+1}^{\infty} \rho^{2k} ,$$

where $C_1 = \|\Sigma\| C^2 d^2 \rho^{2v}$, and $M \geq K$ for some integer K. Then letting $s = k - (M - p + 1)$, we have for $M \geq K$, independent of v,

$$\|R(v) - R_M(v)\| \leq C_2 \rho^{2M} = C_2 \gamma^M .$$

Thus the iterative method consists of the following steps:

(a) Find $B(0), \ldots, B(p)$ and $R_p(0), \ldots, R_p(p)$ by (2.10) and (2.11).

(b) Given $B(M-p), \ldots, B(M)$ and $R_M(0), \ldots, R_M(p)$, find $B(M+1)$ by (2.10) and the $R_{M+1}(\cdot)$ by

$$R_{M+1}(v) = R_M(v) + B(M+1-v)\Sigma B^T(M+1) , \quad v = 0, \ldots, p .$$

(c) Iteration stops if for a given δ,

$$\frac{\|R_{M+1}(v) - R_M(v)\|}{\|R_M(v)\|} < \delta , \qquad v = 0, \ldots, p .$$

If not, return to (b).

If the algorithm does not converge, the approximant is not stationary. If convergence is achieved, Γ_p can be checked for positive definiteness.

Comparison of the Methods

(a) In most cases knowledge of the $R(\cdot)$ is desirable, so methods 2 and 3 seem preferable.

(b) To write a computer program of method 2 requires approximately p^2d^4 storage locations, while method 3 requires $3(p+1)d^2$ locations $(A(1),\ldots,A(p),\Sigma,B(M-p),\ldots,B(M)$, and $R_M(0),\ldots,R_M(p))$.

(c) Preliminary empirical evidence indicates that the value of M necessary to attain convergence in method 3 is not large.

(d) Method 3 does not have the problem of the matrix W of method 2 being ill-conditioned. However, method 3 is an iterative method and may suffer the difficulties of any iterative algorithm.

Thus, unless the values of p and d are quite small, method 3 appears to be the most useful algorithm.

3. Applications of the Algorithms to Obtain $R(\cdot)$

An Application to Scalar Mixed Time Series

Consider the scalar mixed autoregressive moving average process, $\{X(t), t\in Z\}$, of order (p,q), defined by

$$\sum_{j=0}^{p} \alpha(j)X(t-j) = \sum_{k=0}^{q} \beta(k)\epsilon(t-k), \quad t\epsilon Z ,$$

where $\{e(t), t\epsilon Z\}$ is a series of uncorrelated random variables with zero mean and variance σ^2, and the zeros of the complex polynomials

$$g(z) = \sum_{j=0}^{p} \alpha(j)z^j$$

$$h(z) = \sum_{k=0}^{q} \beta(k)z^k$$

are outside the unit circle. The information matrix (Whittle (1953)) of the parameters $\underset{\sim}{\alpha} = (\alpha(1),\ldots,\alpha(p))^T$ and $\underset{\sim}{\beta} = (\beta(1),\ldots,\beta(q))^T$ is given by (see Hannan (1970), p. 392)

$$I = \begin{bmatrix} I^{\alpha\alpha} & I^{\alpha\beta} \\ (I^{\alpha\beta})^T & I^{\beta\beta} \end{bmatrix} ,$$

where

$$(I^{\alpha\alpha})_{jk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i(j-k)\omega}}{|g(e^{i\omega})|^2} \, d\omega , \qquad j,k = 1,\ldots,p ,$$

$$(I^{\beta\beta})_{jk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i(j-k)\omega}}{|h(e^{i\omega})|^2} \, d\omega \, , \qquad j,k = 1,\ldots,q \, ,$$

$$(I^{\alpha\beta})_{jk} = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i(j-k)\omega}}{g(e^{i\omega})h(e^{-i\omega})} \, d\omega \, , \qquad j = 1,\ldots,p \quad k = 1,\ldots,q \, .$$

Consider the two dimensional autoregressive process $\underset{\sim}{u}(\cdot)$ of order $m = \max(p,q)$ defined by

$$G(L)\underset{\sim}{u}(t) = \underset{\sim}{e}(t) \, ,$$

where

$$\underset{\sim}{u}(t) = (u_1(t) \, , \, u_2(t))^T \, , \quad \underset{\sim}{e}(t) = (\varepsilon(t), \varepsilon(t))^T \, ,$$

and

$$G(z) = \sum_{j=0}^{m} \begin{bmatrix} \alpha'(j) & 0 \\ 0 & \beta'(j) \end{bmatrix} z^j$$

$$= \begin{bmatrix} g(z) & 0 \\ 0 & h(z) \end{bmatrix} \, ,$$

where $\alpha'(j) = \alpha(j)$ for $j = 0,\ldots,p$, and zero otherwise, and $\beta'(j) = \beta(j)$ for $j = 0,\ldots,q$, and zero otherwise. Assume $E[\varepsilon^2(t)] = 1$ and thus

$$E[\underset{\sim}{e}(t)\underset{\sim}{e}^T(t)] = \underset{\sim}{1}\,\underset{\sim}{1}^T \ ,$$

where $\underset{\sim}{1}$ is the two dimensional vector of 1's. Then the spectral density $f_{\underset{\sim}{u}}(\cdot)$ of the $\underset{\sim}{u}(\cdot)$ process is given by

$$f_{\underset{\sim}{u}}(\omega) = \frac{1}{2\pi}\, G^{-1}(e^{i\omega}) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} G^{-*}(e^{i\omega}) \ , \quad \omega\in[-\pi,\pi]$$

$$= \frac{1}{2\pi} \begin{bmatrix} \dfrac{1}{|g(e^{i\omega})|^2} & \dfrac{1}{g(e^{i\omega})h(e^{-i\omega})} \\ \dfrac{1}{g(e^{-i\omega})h(e^{i\omega})} & \dfrac{1}{|h(e^{i\omega})|^2} \end{bmatrix} .$$

Defining

$$R_{\underset{\sim}{u};j,k}(v) = \int_{-\pi}^{\pi} f_{\underset{\sim}{u};j,k}(\omega)e^{iv\omega}\, d\omega \ ,$$

where $f_{\underset{\sim}{u};j,k}(\omega)$ denotes the $(j,k)^{th}$ element of $f_{\underset{\sim}{u}}(\omega)$, we have

$$(I^{\alpha\alpha})_{jk} = R_{\underset{\sim}{u};1,1}(j-k) \ , \qquad j,k = 1,\ldots,p \ , \tag{3.1}$$

$$(I^{\beta\beta})_{jk} = R_{\underset{\sim}{u};2,2}(j-k) \ , \qquad j,k = 1,\ldots,q \ , \tag{3.2}$$

$$(I^{\alpha\beta})_{jk} = -R_{u;1,2}(j-k) \ , \qquad j = 1,\ldots,p, \ k = 1,\ldots,q \ . \tag{3.3}$$

Thus to find I in a convenient way, without any numerical quadrature, we can find the $R_{\underset{\sim}{u}}(\cdot)$ by the methods of Section 2, and form I from (3.1), (3.2), (3.3). This is of particular importance when estimating the parameters of the scalar mixed process (see Hannan (1970), p. 390).

An Application to Constrained Autoregressive Estimation

Given a sample $\underset{\sim}{x}(1),\ldots,\underset{\sim}{x}(T)$, consider the problem of efficiently estimating the autocovariance sequence $R(\cdot)$ of a multiple autoregressive process $\underset{\sim}{X}(\cdot)$ where the coefficients $\underset{\sim}{\alpha}_1 = \text{vec}\,(A(1) \vdots \ldots \vdots A(p))$ are subject to the linear constraints

$$B\underset{\sim}{\alpha}_2 = \underset{\sim}{\alpha}_1 \ . \tag{3.4}$$

Hannan and Terrell (1972) consider problems of a similar nature.

Efficient estimates $\hat{A}(1),\ldots,\hat{A}(p)$, and $\hat{\Sigma}$ of $A(1),\ldots,A(p)$ and Σ (ignoring the constraints) are obtained by solving the Yule-Walker equations (2.3) with

$$R_T(v) = \frac{1}{T} \sum_{t=1}^{T-v} \underset{\sim}{x}(t)\underset{\sim}{x}^T(t+v) \ , \qquad v = 0, \ldots, p$$

replacing $R(v)$. Then an estimator $\hat{I}^{\alpha_2\alpha_2}$ of the information matrix $I^{\alpha_2\alpha_2}$ of $\underset{\sim}{\alpha}_2$ is given by (see Hannan (1970), p. 329)

$$\hat{I}^{\alpha_2\alpha_2} = \hat{\Gamma}_p \otimes \hat{\Sigma}^{-1}$$

where $\hat{\Gamma}_p = \text{TOEPL}\{R_T(0), \ldots, R_T(1-p)\}$ and
$C \otimes D$ denotes the Kronecker product of the $(n \times m)$ matrix $C = (C_{jk})$ and the $(r \times s)$ matrix $D = (D_{jk})$:

$$C \otimes D = \begin{bmatrix} C_{11}D & \cdots & C_{1m}D \\ \vdots & & \vdots \\ C_{n1}D & \cdots & C_{nm}D \end{bmatrix} .$$

Then an efficient estimator $\hat{\underset{\sim}{\alpha}}_1$ of $\underset{\sim}{\alpha}_1$ is given, by applying the theory of optimization under linear constraints (Pagano (1974)), as the solution of the normal equations

$$(B^T \hat{I}^{\alpha_2\alpha_2} B)\hat{\underset{\sim}{\alpha}}_1 = B^T \hat{I}^{\alpha_2\alpha_2} \hat{\underset{\sim}{\alpha}}_2 \ .$$

Then to obtain efficient estimators $\hat{R}(1), \ldots, \hat{R}(p)$ of $R(0), \ldots, R(p)$ subject to (3.4), one can use the methods of Section 2.

4. Numerical Example

Let $p = 2$, $d = 2$, and

$$A(1) = \begin{bmatrix} -.2580 & -.1429 \\ .5572 & -.8637 \end{bmatrix}, \quad A(2) = \begin{bmatrix} -.1471 & .1280 \\ -.4732 & .3305 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} 2.6034 & .9053 \\ .9053 & 2.1450 \end{bmatrix}.$$

Method 2 gives the exact values of $R(\cdot)$

$$R(0) = \begin{bmatrix} 3.0 & 1.0 \\ 1.0 & 4.0 \end{bmatrix}, \quad R(1) = \begin{bmatrix} 1.0 & -.50 \\ .50 & 2.0 \end{bmatrix}$$

$$R(2) = \begin{bmatrix} .50 & .10 \\ .05 & .60 \end{bmatrix}.$$

After 9 iterations, method 3 gives values, correct to $\delta = 1 \times 10^{-5}$

$$R(0) = \begin{bmatrix} 2.9999974 & 1.0000470 \\ 1.0000470 & 4.0000683 \end{bmatrix}$$

$$R(1) = \begin{bmatrix} .9999955 & -.4999354 \\ .5000751 & 2.0000473 \end{bmatrix}$$

$$R(2) = \begin{bmatrix} .4998513 & .1000909 \\ .0499245 & .5999987 \end{bmatrix} .$$

Thus for this second order autoregressive process, one needs an approximating moving average process of order 11 to obtain five place accuracy. (Calculations performed on a CDC-6400.)

References

Akaike, H. (1971), "Autoregressive model fitting for control," Ann. Inst. Statist. Math., 23, 163-180.

Hannan, E. J. (1970), Multiple Time Series, New York, John Wiley.

Hannan, E. J. and Terrell, R. D. (1972), "Time series regression with linear constraints," International Economic Review, 13, 189-200.

Masani, P. (1966), "Recent Trends in Multivariate Prediction Theory," Multivariate Analysis, P. R. Krishnaiah, Ed., New York: Academic Press, 351-382.

Pagano, M. (1973), "When is an autoregressive scheme stationary?," Communications in Statistics, 1, 533-544.

Pagano, M. (1974), "Modification of parameter estimates under linear constraints," Technical Report No. 12, Statistical Science Division, State University of New York at Buffalo.

Parzen, E. (1974), "Some recent advances in time series modeling," IEEE Transactions on Automatic Control, AC-19, 723-730.

Parzen, E. (1976), "Multiple time series: determining the order of approximating autoregressive schemes," Multivariate Analysis, 6.

Quenouille, M. H. (1957), The Analysis of Multiple Time Series, New York Hafner Publishing Company.

Robinson, E. A. (1967), Multichannel Time Series Analysis with Digital Computers Programs, San Francisco, Holden-Day.

Schur, I. (1917), "Uber potenzreihen, die in innern des Einheitskreises beschränkt sind," J. Reine Angew. Math., 147, 205-232.

Whittle, P. (1953), "The analysis of multiple stationary time series," Journal of the Royal Statistical Society, B, 15, 125-139.

Whittle, P. (1963), "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," Biometrika, 50, 129-134.

Wilkinson, J. H. (1967), "The solution of ill-conditioned linear equations," in Mathematical Methods for Digital Computers, Eds. A. Ralston and H. S. Wilf, New York, John Wiley.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. 43
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): On the Stationarity of Multiple Autoregressive Approximants: Theory and Algorithms
5. TYPE OF REPORT & PERIOD COVERED: Technical
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): H. Joseph Newton and Marcello Pagano
8. CONTRACT OR GRANT NUMBER(s): N00014-72-C-0508
9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Statistical Laboratory
State University of New York at Buffalo
Amherst, New York 14226
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-234
11. CONTROLLING OFFICE NAME AND ADDRESS:
Office of Naval Research
Statistics and Probability Program Code 436
Arlington, Virginia 22217
12. REPORT DATE: Aug[redacted]76
13. NUMBER OF PAGES: 26
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 27 P.
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; Distribution Unlimited

18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiple autoregression
Stability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Numerical methods are presented for determining the stability of a multiple autoregressive scheme. The question of stability is important in prediction theory, but, furthermore, the methods for determining stability are such that they yield information which is important when fitting time series models to data.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

401 511